

## МЕТОД ФОРМУВАННЯ КОНТЕНТУ В СИСТЕМАХ ЕЛЕКТРОННОЇ КОНТЕНТ-КОМЕРЦІЇ

© Берко А., Висоцька В., Чирун Л., 2013

Запропоновано метод формування контенту як етап життєвого циклу контенту в системах електронної комерції. Метод реалізує процеси опрацювання інформаційних ресурсів у системах електронної контент-комерції та спрощує технологію автоматизації формування комерційного контенту. Проаналізовано основні проблеми електронної контент-комерції та функціональних сервісів формування комерційного контенту. Запропонований метод дає можливість створити засоби опрацювання інформаційних ресурсів та реалізувати модуль формування контенту.

**Ключові слова:** інформаційний ресурс, комерційний контент, контент-аналіз, контент-моніторинг, контентний пошук, система електронної контент-комерції.

**In the given article content is forming method as the content life cycle stage in electronic commerce systems is proposed. The method implements the information resources processing in electronic content commerce systems and automation technology simplifies the commercial content formation. In the given article the main problems of electronic content commerce and functional services of commercial content forming are analyzed. The proposed method gives an opportunity to create an instrument of information resources processing and to implement the module of content forming.**

**Key words:** information resources, commercial content, content analysis, content monitoring, content search, electronic content commerce systems.

### Вступ. Загальна постановка проблеми

Активний розвиток Інтернету сприяє зростанню потреб у даних виробничого/стратегічного характеру і реалізації нових форм інформаційного обслуговування [6, 9]. Документована інформація, підготовлена відповідно до потреб користувачів і призначена для їх задоволення, є інформаційним продуктом або комерційним контентом. Розроблення і впровадження систем електронної контент-комерції є одним із стратегічних напрямів розвитку електронного бізнесу. Характерною рисою таких систем є можливість автоматичного опрацювання інформаційних ресурсів для збільшення обсягів продажу контенту постійному користувачу, активного залучення потенційних користувачів та розширення меж цільової аудиторії [6]. Проблематика проектування, створення, впровадження та супроводу систем електронної контент-комерції є актуальною з огляду активного розвитку досліджень у галузі електронного бізнесу. Важливою проблемою залишається відсутність теоретичного обґрунтування, стандартизованих методів і потреба в уніфікації програмних засобів опрацювання інформаційних ресурсів у таких системах. З'являються нові підходи та способи вирішення цієї проблеми, але важливим питанням залишається невідповідність між відомими методами і засобами опрацювання інформаційних ресурсів та принципами побудови систем електронної контент-комерції. Відсутні загальні підходи до створення систем електронної контент-комерції та відповідні методи опрацювання інформаційних ресурсів у цих системах.

### Аналіз останніх досліджень та публікацій

Однією із головних особливостей сучасності є постійний ріст темпів виробництва контенту. Цей процес є об'єктивним і позитивним, але виникла проблема: прогрес у галузі виробництва контенту призводить до пониження загального рівня інформованості потенційного користувача. Крім збільшення

обсягів контенту до масштабів, яке призводить до неможливості його безпосереднього опрацювання та швидкості його поширення, виникла низка специфічних проблем (табл. 1) [7–10].

Таблиця 1

**Основні проблеми формування комерційного контенту, розроблено за [9]**

Назва	Основна причина	Рішення
Інформаційний шум	Структурованість масивів контенту.	Фільтри, контент-моніторинг, аналіз сайту, контент-аналіз.
Паразитичний контент	Поява в якості додатків.	Фільтри, контент-моніторинг, контент-аналіз.
Нерелевантність контенту	Невідповідність потребам користувачів.	Створення анотованої бази даних, пошукових образів первинного контенту та їх кластеризація, контент-аналіз.
Дублювання контенту	Дублювання в джерелах.	Контент-аналіз, сканери і фільтри на базі статистики та критеріїв.
Навігація в потоці контенту	Швидкий ріст обсягу і поширення контенту.	Аналіз сайту, фільтри, контент-моніторинг, контент-аналіз.
Надмірність результату пошуку	Дублювання і нерелевантність.	Анотований пошук, контент-аналіз та реферування.

Причина втрати актуальності традиційних інформаційно-пошукових систем полягає у швидкому рості обсягу, актуальності та нерівномірній динаміці потоків контенту (постійне систематичне та нерегулярне оновлення контенту). Охоплення та узагальнення великих динамічних потоків контенту вимагає якісно нових методів/підходів для розв'язання задач створення та опрацювання контенту [7–10]. Застосування засобів контент-моніторингу дає можливість автоматизувати знаходження найважливіших складових у потоках/джерелах контенту. Їх застосування викликане потребами систематичного відстеження тенденцій та процесів у середовищі контенту, що постійно оновлюється. Контент-моніторинг – це змістовний аналіз потоків контенту з метою постійного отримання необхідних якісних та кількісних зрізів контенту протягом наперед не визначеного проміжку часу [9, 12, 14]. Складовою контент-моніторингу є контентний пошук та контент-аналіз [1, 4, 5, 7–11, 15, 16]. Контентний пошук – це сукупність операцій, необхідних для знаходження в наперед визначених джерелах контенту, який відповідає запиту користувача природною мовою [8, 9, 12, 14]. Робота з текстом природною мовою є складним завданням для математичної лінгвістики. Із вирішенням задачі морфологічного аналізу робота з текстом реалізується в межах речень або тексту природною мовою як набором лінійно впорядкованих слів, словосполучень або речень.

Для автоматизованого контентного пошуку та опрацювання текстового контенту великого значення набуває наявність/відсутність та частота появи тієї чи іншої категорії лінгвістичної одиниці в досліджуваному масиві контенту [15, 16]. Кількісний підрахунок дає змогу зробити об'єктивні висновки щодо спрямованості контенту за кількістю уживань одиниць аналізу (ключових цитат) у досліджуваних масивах, наприклад, кількість позитивних/негативних відгуків на певний вид продукції [9, 15, 16]. Якісний аналіз дає змогу зробити об'єктивні висновки щодо наявності в масиві контенту шуканої лінгвістичної одиниці та напряму її контексту [9, 15, 16]. Контентний пошук виконують не за текстом контенту, а за його стислими характеристиками – пошуковими образами (ПОб), де основний текст контенту подають у термінах спеціалізованої інформаційно-пошукової мови [8, 12, 14]. Процедура визначення ПОб передбачає індексування, семантичний аналіз основного тексту контенту і переклад його інформаційно-пошуковою мовою (табл. 2) [8, 12, 14]. У модулі зберігають не тексти контенту, а його ПОб. Для пошуку проіндексованого контенту застосовують контент-аналіз до інформаційних запитів. Інформаційний запит, перекладений інформаційно-пошуковою мовою і доповнений для пошуку додатковими даними, є пошуковим розпорядженням (ПР) [8, 12, 14]. Ступінь докладності подання контенту в ПОб його центральній темі/предметі, а також супутніх тем/предметів є глибиною індексування. Автоматизація цього процесу дає змогу забезпечити його уніфікацію, звільнивши частину персоналу від непродуктивної праці з індексування контенту [12–16]. Контентний пошук забезпечують набір

семантичних засобів: інформаційно-пошукова мова, методи індексування контенту/запитів та пошуку [12–16]. Основу семантичних засобів становить інформаційно-пошукова мова – спеціалізована штучна мова, яка призначена для опису центральних тем/предметів і формальних характеристик контенту, а також для опису запитів і виконання пошуку [12–16]. На практиці одну мову використовують для індексації контенту, а іншу – для індексації інформаційних запитів.

Таблиця 2

**Основні етапи операції контентного пошуку, розроблено за [8, 12, 14]**

Назва операції	Характеристика операції
Формування ПОБ	Створення, введення, зберігання в модулі ПОБ.
Формування запитів і ПР	Створення, введення та зберігання в модулі запитів і ПР запиту користувача.
Пошук контенту	Порівняння ПОБ контенту з ПР запиту користувача.
Контент-аналіз	Кількісний та якісний аналіз текстового контенту.
Формування результату	Результат застосування контент-аналізу контенту позитивний в межах (0,7; 1] або (0,5; 1].
Прийняття рішення	Прийняття рішення про видавання контенту відповідно до результату застосування контент-аналізу.
Подання контенту	Видавання контенту, що відповідає інформаційному запиту користувача.

Форматування контенту – це процес індексування, семантичного аналізу, визначення основного змісту контенту та перетворення його на XML-формат. Форматування контенту виконують вручну модератором або автоматично засобами контент-аналізу [9]. Під час індексування досліджують текст контенту, визначають його центральну тему та описують його в термінах інформаційно-пошукової мови [12–16]. У контенті назви розділів, як правило, розкривають центральну тему і предмет, але за назвою не завжди можна ідентифікувати контент. Природну мову не використовують як інформаційно-пошукову через численні граматичні вclusions, недостатнє структурування, неоднозначність і велику надлишковість, зокрема, для української мови 75–80 %. В інформаційно-пошуковій мові серед основних елементів (табл. 3) не використовують синоніми та омоніми через їх семантичну неоднозначність [1–5, 11–16].

Таблиця 3

**Основні елементи інформаційно-пошукової мови, розроблено за [2–3, 12–14]**

Назва елементу	Характеристика елементу мови
Алфавіт	Множина графічних знаків для фіксації слів і висловлювань мови.
Лексика (парадигматика)	Множина пов'язаних між собою мовних одиниць, тобто використовуваних у мові слів.
Грамматика (синтагматика)	Множина правил об'єднання мовних одиниць у словосполучення, тобто дієвих засобів побудови речень.
Парадигми	Лексико-семантичні групи слів з предметно-логічними зв'язками на основі семантичних ознак.
Парадигматичні відношення	Базові та аналітичні відношення між словами, що не залежать від контексту, в якому вони використовуються, і породжені не мовними, а логічними зв'язками.
Синтагматичні відношення	Лінійні відношення між словами, які встановлюються при об'єднанні слів у словосполучення та фрази.
Правила ідентифікації індексів	Парадигматика (лексика) та синтагматика (грамматика) мови.
Висловлювання	Реалізоване речення, висловлювання є реченням природною мовою, проте зворотне не вірно.
Міжфразова єдність	Множина висловлювань, об'єднаних семантично і синтаксично в фрагмент. Ядром міжфразової єдності є висловлення, яке не підпорядковане іншим висловлюванням і зберігає сенс при виділенні з контексту.
Фрагменти-блоки	Множина міжфразових єдностей, які забезпечують цілісність тексту через змістові та тематичні зв'язки.

Доцільність використання інформаційно-пошукових мов залежить від призначення засобів пошуку, рівня технічних засобів, автоматизації інформаційних процедур і ланки управління [12–16]. При розробленні інформаційно-пошукових мов звертають увагу на такі моменти: специфіка галузі/теми, для якої розробляється мова; особливості текстів з пошукового масиву контенту; характер інформаційних потреб користувачів системи електронної контент-комерції [12–13].

### **Виділення проблем**

Текстовий контент (стаття, коментар, книга тощо) містить значний обсяг тексту природною мовою, де частина інформації є абстрактною. Текст – це об'єднана за змістом послідовність лексичних одиниць, основними властивостями якої є інформаційна, структурна, комунікативна зв'язність і цілісність, що відображають змістовну і структурну сутність тексту [3]. Як функціонально-семантико-структурна єдність текст володіє правилами побудови, виявляє закономірності змістовного та формального з'єднання складових одиниць [2, 3, 13]. Зв'язність тексту визначається через зовнішні структурні показники та формальну залежність компонентів тексту, а цілісність – через тематичну, концептуальну та модальну залежність [3]. Текст реалізує структурно подану діяльність, що передбачає суб'єкт і об'єкт, процес, мету, засоби і результат, які відображаються в змістовно-структурних, функціональних, комунікативних показниках тексту. Семантика тексту зумовлена комунікативним завданням передавання контенту [13]. Зовнішня (композиційна) і внутрішня структура тексту визначається внутрішньою організацією одиниць тексту і закономірностями їх взаємозв'язку в тексті як цілісного контенту. На композиційному рівні виділяють речення, абзаци, параграфи, розділи, розділи, підрозділи, сторінки тощо, які, крім речення, побічно пов'язані з внутрішньою структурою, тому не розглядаються [2, 3, 13]. При використанні статистичних методів аналізу ігнорують лінгвістичну взаємопов'язаність і нелінійність природної мови. Незадіяність проміжних рівнів подання тексту у вигляді семантичних структур пояснюють відсутністю ефективного формалізму опису структури тексту. Одиниці семантико-граматичного (синтаксичного) та композиційного рівня тексту знаходяться у взаємозв'язку і взаємозумовленості, в окремому випадку збігаються, накладаючись один на одного (наприклад, міжфразова єдність і абзац, хоча при цьому вони зберігають відмінні ознаки). Із семантичною, граматичною та композиційною структурою тексту пов'язані його стильові і стилістичні характеристики. Кожен текст виявляє функціонально-стильову орієнтацію (наукову, художню тощо) і володіє стилістичними якостями, залежними від індивідуальності автора та орієнтації тексту [2-3, 13]. Це ускладнює процес форматування комерційного контенту різних авторів.

### **Формулювання мети**

Метою статті є розроблення методу формування комерційного контенту для опрацювання інформаційних ресурсів в системах електронної контент-комерції. Актуальність роботи полягає у необхідності отримувати оперативні/об'єктивні оцінки рівня конкуренції на сегменті фінансового ринку комерційного контенту; оцінювати рівень конкурентів та міри їх конкурентоспроможності на фінансовому ринку з поширення контенту. З позиції системного підходу необхідно дослідити етапи опрацювання інформаційних ресурсів та розробити оптимальний життєвий цикл для процесу формування контенту. Розроблення методу формування контенту дає можливість створити засоби опрацювання інформаційних ресурсів та автоматичного формування контенту.

### **Аналіз отриманих наукових результатів**

Процес формування комерційного контенту для інформаційного ресурсу забезпечує зв'язок між множиною вхідних даних із різних джерел інформації та множиною сформованого комерційного контенту, збереженого у відповідній базі даних у системах електронної контент-комерції, що можна подати як  $Source(x_i) \rightarrow x_i \rightarrow X \rightarrow Formation(u_f, x_i, t_p) \rightarrow c_r \rightarrow C \rightarrow DataBase(C)$ , де  $Source(x_i)$  – джерело  $x_i$  контенту,  $x_i$  –  $i$ -ий контент з джерела;  $X$  – множина даних з відповідного джерела;  $Formation(u_f, x_i, t_p)$  – оператор формування контенту з  $x_i$  контенту джерела в фіксований час  $t_p$  за умов формування контенту  $u_f$ ;  $c_r$  – сформований за умови  $u_f$   $r$ -й

контент;  $C$  – множина сформованого контенту;  $DataBase(C)$  – база даних сформованого комерційного контенту.

Модель процесу формування контенту в системах електронної контент-комерції подамо як

$$Formation = \left\langle \begin{array}{l} X, Gathering, Formatting, KeyWords, Backup, \\ Categorization, BuDigest, Dissemination, T, C \end{array} \right\rangle, \quad (1)$$

де  $X = \{x_1, x_2, \dots, x_{n_x}\}$  – множина вхідних даних  $x_i \in X$  з різних інформаційних ресурсів або від модераторів при  $i = \overline{1, n_x}$ ;  $Gathering$  – оператор збирання/створення контенту з різних джерел;  $Formatting$  – оператор форматування контенту;  $KeyWords$  – оператор виявлення ключових слів і понять контенту;  $Categorization$  – оператор рубрикації контенту;  $Backup$  – оператор виявлення дублювання контенту;  $BuDigest$  – оператор формування дайджестів контенту;  $Dissemination$  – оператор вибіркового поширення контенту;  $T = \{t_1, t_2, \dots, t_{n_t}\}$  – час  $t_p \in T$  транзакції формування контенту при  $p = \overline{1, n_t}$ ;  $C = \{c_1, c_2, \dots, c_{n_c}\}$  – множина комерційного контенту  $c_r \in C$  при  $r = \overline{1, n_c}$ .

Процес формування контенту описано оператором вигляду  $c_r = Formation(u_f, x_i, t_p)$ , де  $u_f$  – множина умов формування контенту, тобто  $u_f = \{u_1(x_i), \dots, u_{n_u}(x_i)\}$ .

Комерційний контент подано так:

$$c_r = \left\{ \bigcup_f u_f \mid (x_i \in X) \wedge (\exists u_f \in U), U = U_{x_i} \vee U_{-x_i}, i = \overline{1, m}, f = \overline{1, n} \right\}, \quad (2)$$

проходячи наступні етапи перетворення з множини даних у множину релевантного, форматowanego, класифікованого та валідованого контенту:

$$x_i \in X \rightarrow Gathering(u_f, x_i, t_p) \rightarrow Backup(c_r, u_b, t_p) \rightarrow Formatting(c_r, t_p) \rightarrow KeyWords(c_r, t_p) \rightarrow \\ Categorization(c_r, t_p) \rightarrow BuDigest(c_r, t_p) \rightarrow Dissemination(c_r, t_p) \rightarrow c_r \in C.$$

Рішення, здатне допомогти орієнтуватися в динамічній вхідній інформації з різних джерел, надають процеси синдикації даних  $C = Gathering(X, U_G, T)$ , тобто збирання інформації з джерел і подальше поширення її фрагментів відповідно до потреб користувачів, де  $X$  – множина контенту з різних джерел даних,  $U_G$  – множина умов збирання даних з різноманітних джерел,  $Gathering$  – оператор збирання/створення контенту,  $T$  – час збирання/створення контенту.

Виявлення дублювання контенту позначено оператором  $C = Backup(Gathering(X, U_G, T), U_B)$ , де  $X$  – множина контенту з різних джерел даних,  $U_B$  – множина умов виявлення дублювання змісту контенту,  $Backup$  – оператор виявлення дублювання змісту контенту,  $C$  – множина контенту. Дубльований за змістом контент виявляють за лінгвостатистичними методами виявлення загальних термів, ланцюжки яких утворюють словесні сигнатури контенту.

Технологія синдикації контенту передбачає процес навчання програм збирання даних із структурними особливостями з окремих джерел (з інформаційних ресурсів, від модераторів, користувачів, відвідувачів, журналістів, редакторів), безпосереднє сканування контенту та приведення до загального:

$$C = Formatting(Backup(Gathering(X, U_G, T), U_B), U_{FR}), \quad (3)$$

де  $Formatting$  – оператор форматування контенту,  $U_G$  – множина умов збирання даних з різноманітних джерел,  $Gathering$  – оператор збирання/створення контенту,  $U_{FR}$  – множина умов форматування інформації,  $T$  – час збирання контенту.

Опрацювання множини контенту  $C$  для виявлення ключових слів побудоване на принципі знаходження ключових слів за змістом (термами), базується на законі Зіпфа і зводиться до вибору слів із середньою частотою появи (найвживаніші слова ігнорують за допомогою стоп-словника, а рідкісні слова з текстів повідомлень не враховують). Виявлення ключових слів та понять визначається оператором  $KeyWords(C)$  та описано оператором вигляду:

$$C = KeyWords(Formatting(Backup(Gathering(X, U_G, T), U_B), U_{FR}), U_K), \quad (4)$$

де  $KeyWords$  – оператор виявлення ключових слів і понять контенту, що реалізується як набір процесів, поданих діаграмою використання на рис. 1;  $Formatting$  – оператор форматування контенту,  $U_G$  – множина умов збирання даних з різних джерел,  $Gathering$  – оператор збирання/створення контенту,  $U_{FR}$  – множина умов форматування даних,  $T$  – час збирання контенту,  $U_K$  – множина умов виявлення ключових слів та понять.

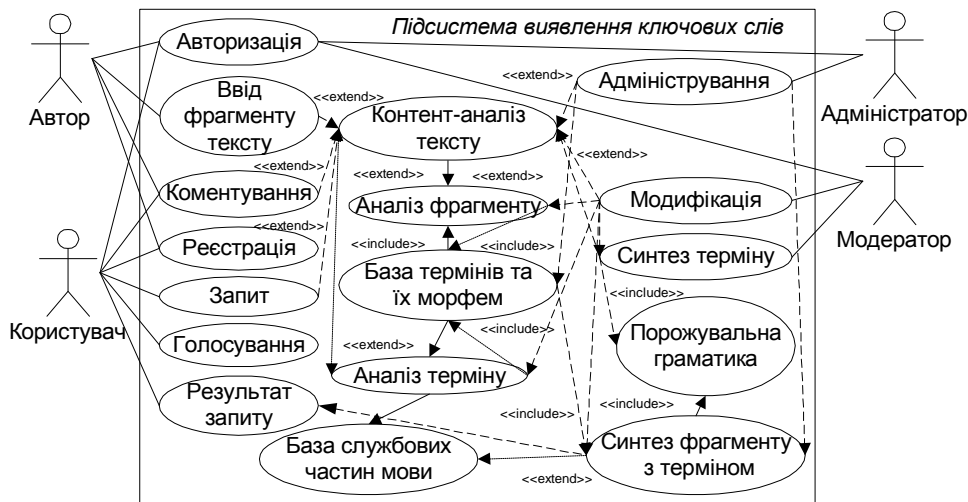


Рис. 1. Діаграма варіантів використання для процесу виявлення ключових слів тематики

За допомогою бази даних термінів/морфем і службових частин мови та визначених правил аналізу тексту шукають термін. За правилами породжувальної граматики коригують термін згідно з правилами його вживання у контексті. На рис. 2 подано діаграму послідовності для процесу виявлення ключових слів тематики контенту.

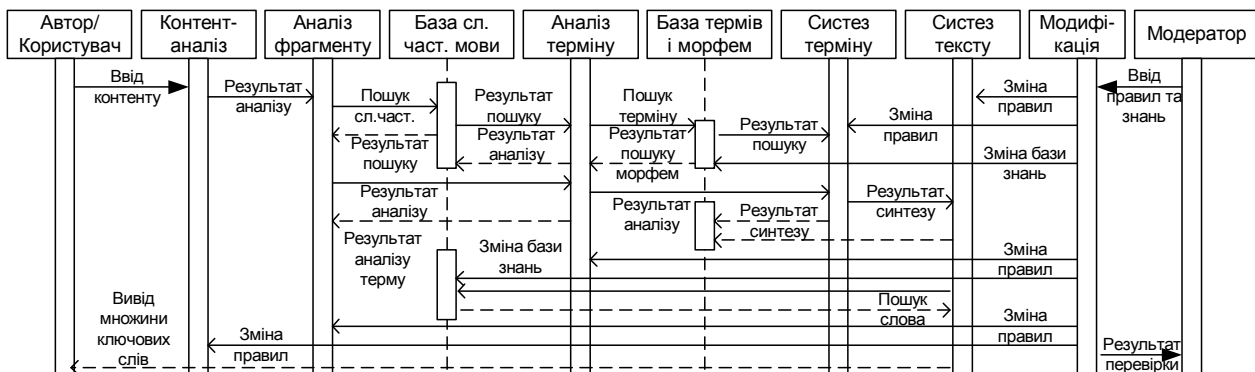


Рис. 2. Діаграма послідовності для процесу виявлення ключових слів тематики контенту

Засоби класифікації та розподілу контенту являють собою інформаційно-пошукову систему вибіркового поширення контенту (контентний роутер). Контент аналізується на відповідність тематичним запитам оператором  $C_{Cl} = Categorization(KeyWords(C, U_K), U_{Cl})$ , де  $KeyWords(C, U_K)$  –

оператор виявлення ключових слів; *Categorization* – оператор рубрикації контенту згідно із виявленими ключовими словами;  $U_K$  – множина умов виявлення ключових слів;  $U_{Ct}$  – множина умов рубрикації;  $C_{Ct}$  – множина рубрикованого релевантного контенту. Множина дайджестів  $C_D$  формується за допомогою такої залежності  $C_D = BuDigest(C_{Ct}, U_D)$ , де *BuDigest* – оператор формування дайджестів;  $U_D$  – множина умов формування дайджестів;  $C_{Ct}$  – множина рубрикованого релевантного контенту, тобто  $C_D = BuDigest(Categorization(KeyWords(C, U_K), U_{Ct}), U_D)$ . Контент розсилається користувачам та завантажується в тематичні бази даних. Вибіркове поширення контенту описується як  $C_{Ds} = Dissemination(C_D, U_{Ds})$ , де  $C_{Ds}$  – множина вибірково поширеного контенту,  $U_{Ds}$  – множина умов вибіркового поширення контенту, *Dissemination* – оператор вибіркового поширення контенту. На рис. 3, а подано діаграму кооперації процесу виявлення ключових слів тематики контенту.

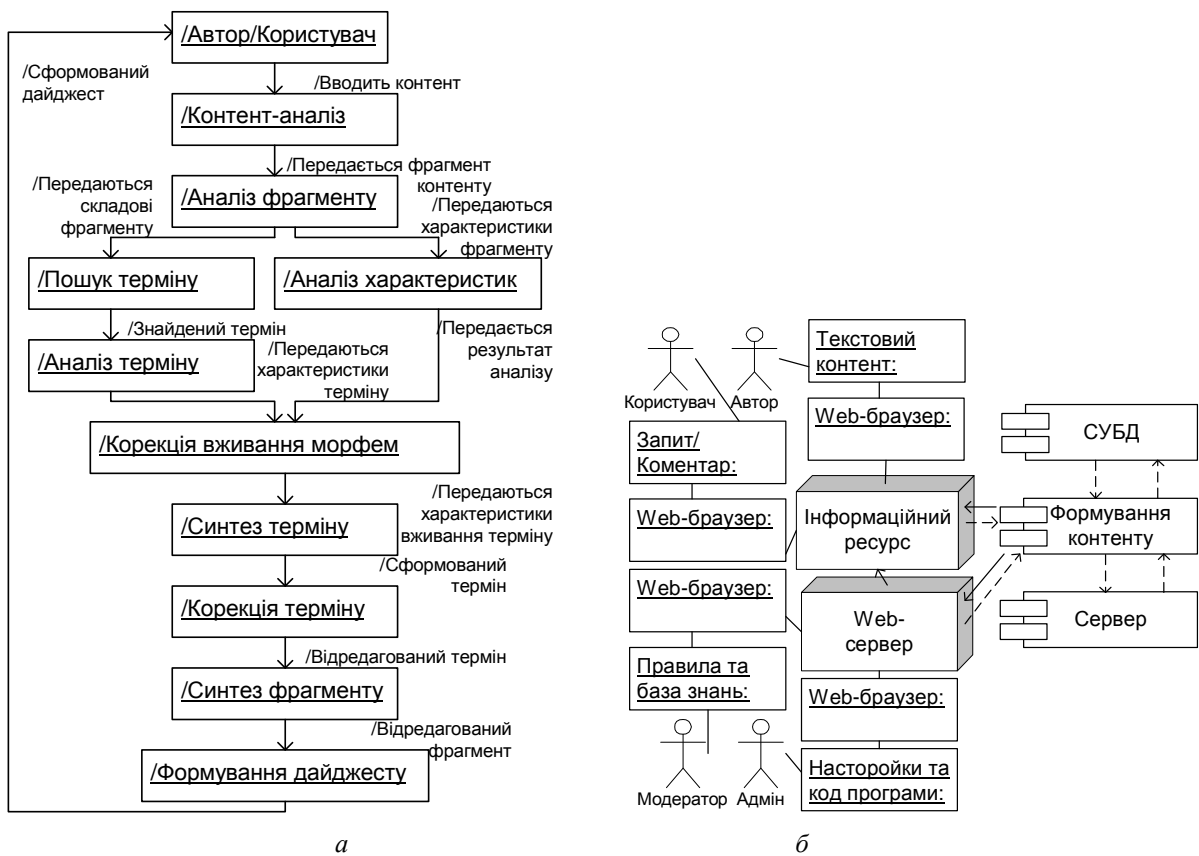


Рис. 3. Діаграма кооперації (а) та компонентів процесу виявлення ключових слів тематики контенту (б)

Аналіз лексико-граматичної та семантико-прагматичної побудови тексту використовують в процесі автоматичної рубрикації контенту, головним завданням якого є знаходження в потоці контенту шляхом контент-аналізу тексту, який найбільше відповідає тематикам контенту та потребам користувачів. Після аналізу фрагмента тексту і терміна синтезується новий термін як ключове слово тематики контенту. На рис. 3, б подано діаграму компонентів процесу виявлення ключових слів тематики контенту. Принцип виявлення ключових слів за змістом (термами) ґрунтується на законі Зіпфа і зводиться до вибору слів із середньою частотою вживання (найвживаніші слова ігнорують за допомогою словника, а рідкісні слова з тексту не враховують) з використанням бази термінів та їхніх морфем. На рис. 4 наведено діаграму діяльності процесу виявлення ключових слів тематики контенту. Наступним кроком формування контенту за цим методом є рубрикація контенту.

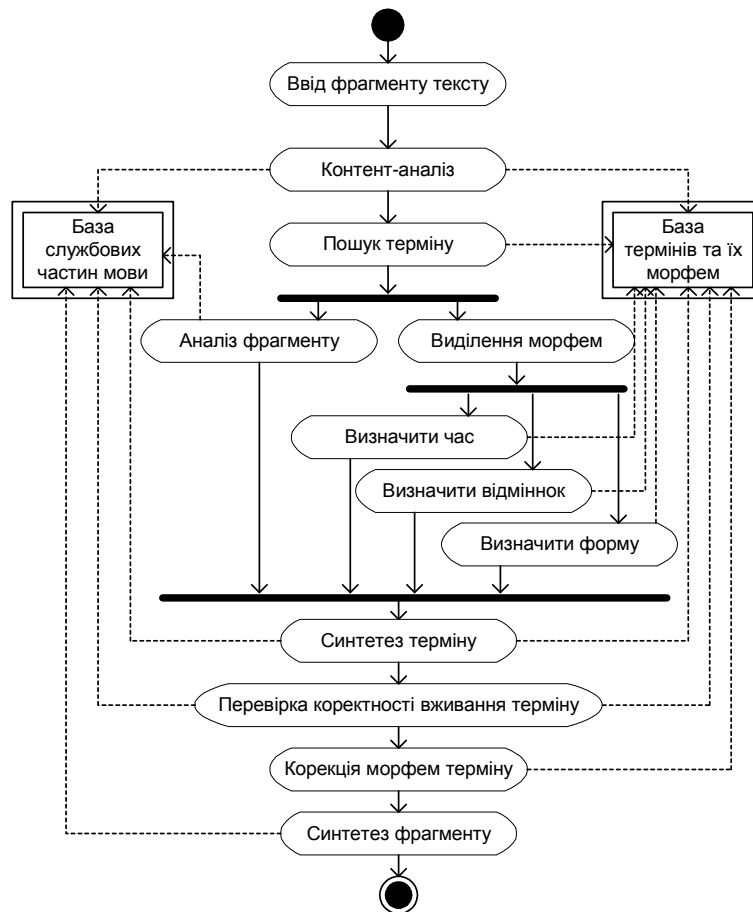


Рис. 4. Діаграма діяльності для процесу виявлення ключових слів тематики контенту

На основі методу розроблено підсистеми формування контенту, які на різних стадіях реалізовано в інтернет-проектах “Фотогалерея Висоцьких” (fotogalereja-vysocjkykh.com), “Татьяна” (tatjana.in.ua), “Вголос” (vgolos.com.ua), “Прес-Тайм” (presstime.com.ua), “AutoChip” (autochip.vn.ua), “Курси валют” (kursyvalyut.com), “Добрий ранок, бухгалтер” (dobryjranok.com). У табл. 4 подано порівняльні характеристики роботи розроблених систем, отримані з Google Analytics (accounts.google.com).

Таблиця 4

**Порівняльні характеристики результатів роботи систем за період часу з 10.2012 р. по 11.2012 р.**

Характеристика систем	Фотогалерея	Вголос	Татьяна	Прес-Тайм	AutoChip	Курси валют	Добрий ранок
Підсистема формування контенту	+/-	+	-	+/-	+	+	+/-
Відвідування	73	326 940	49	167 856	406	103	58
Унікальні відвідувачі	62	217 719	21	123 756	326	42	7
Перегляд сторінок	136	562 455	142	345 234	863	237	226
Сторінок/відвідування	1,86	1,72	2,90	1,45	2,13	1,67	3,90
Середня тривалість відвідування (хв:с)	00:47	01:45	04:38	01:09	01:08	00:37	09:35
Показник відмовлень (%)	71,23	76,92	46,94	79,56	56,90	61,23	48,28
Нові відвідування (%)	80,82	51,83	36,43	45,65	77,59	90,87	12,07
Returning Visitor (%)	82,19	48,15	63,27	54,35	77,59	62,79	87,93
New Visitor (%)	17,81	51,85	36,43	45,65	22,41	37,21	12,07
Відвідувачі з України (%)	87,67	89,81	71,43	92,33	73,89	97,07	55,17
Відвідувачі з Росії (%)	2,74	2,55	24,49	6,27	17,00	1,05	43,10
Відвідувачі з США (%)	1,37	0,58	0,07	0,06	0,05	0,61	1,72
Пошуковий трафік (%)	69,86	36,03	73,47	60,05	88,67	59,03	43,10
Трафік переходів (%)	12,33	54,62	0	34,65	3,45	35,65	6,90
Прямий трафік (%)	17,82	9,21	26,53	5,25	7,88	5,32	50,00
Трафік кампаній (%)	0	0,14	0	0,05	0	0	0



Google Analytics забезпечує розширені можливості аналізу даних та дає змогу оцінити трафік контенту та ефективність маркетингових заходів, наприклад, для газети Вголос (рис. 5).

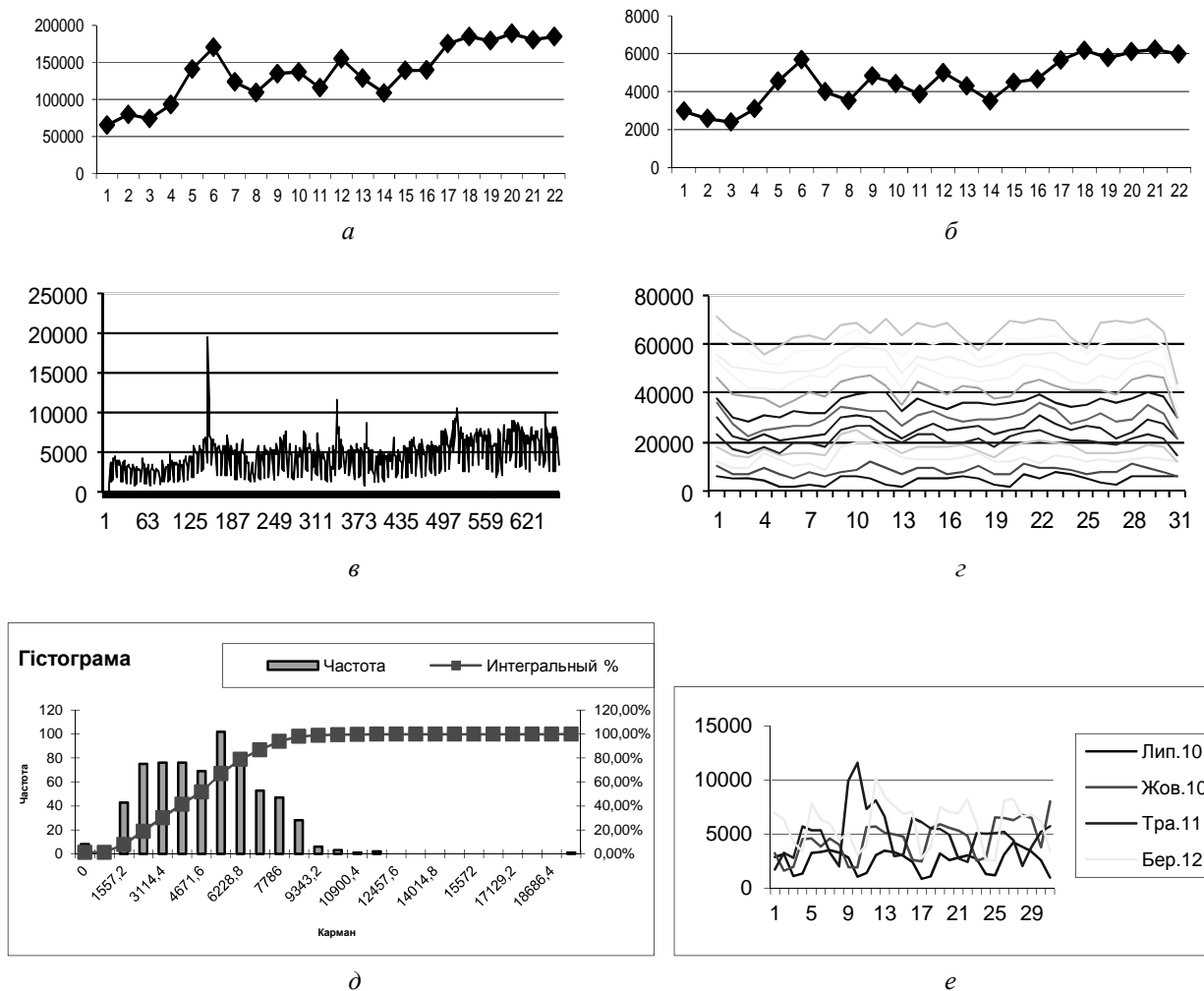


Рис. 5. Розподіл загальної (а), середньої (б), щоденної (в), кількості щомісячних відвідувачів (z) за 2010–2012 рр.

Модель формування комерційного контенту реалізують у вигляді комплексів контент-моніторингу зі збирання контенту з різних джерел даних, і забезпечують створення бази даних контенту відповідно до інформаційних потреб користувачів. У результаті збирання і первинного опрацювання контент приводять до єдиного формату, класифікують відповідно до визначеного рубрикатора та йому приписують дескриптори з ключовими словами. Це полегшує реалізацію процесу управління комерційним контентом. При аналізі тексту досліджують його багаторівневу структуру: вихідний текст як лінійна послідовність символів; лінійна послідовність морфологічних структур; лінійна послідовність висловлювань; мережа взаємопов'язаних міжфразових єдностей. Попереднє опрацювання тексту передбачає поділ тексту на окремі лексеми, який здійснюють методом скінченних автоматів. Вхідною інформацією є текст природною мовою як послідовність символів, а вихідною інформацією – таблиці розділів, речень і лексем аналізованого тексту. Існує така залежність: що більше унікального контенту в системі електронної контент-комерції, то більше відвідувачів інформаційного ресурсу цієї системи. Підсистема формування комерційного контенту скорочує час для заповнення інформаційного ресурсу унікальним контентом та сприяє збільшенню обсягів за короткий проміжок часу унікального контенту на даному інформаційному ресурсі та кількості запитів з пошукових систем. Отримані дані враховують, створюючи або оновлюючи інформаційний ресурс та удосконалюючи архітектуру систем електронної контент-комерції.

## Висновки і перспективи подальших наукових розвідок

Описано метод формування комерційного контенту, який ґрунтується на багаторівневій моделі організації процесів. Така модель передбачає поділ загального процесу на такі етапи: збирання/створення контенту з різних джерел, форматування, виявлення ключових слів і понять, рубрикації, виявлення дублювання, формування дайджестів та вибіркового поширення контенту між модераторами та користувачами системи електронної контент-комерції. В основу методу покладено принципи контент-аналізу, що дає змогу автоматизувати різні етапи створення інформаційного продукту такого типу без втрати змісту і зниження якості. Ефективність методу підтверджують результати його застосування при розробленні низки комерційних контент-проектів. Розроблені засоби автоматизації формування контенту дають змогу прискорити процес формування контенту та підвищити рейтинг використання сформованих за їх допомогою комерційних інформаційних ресурсів.

1. Аверьянов Л. Контент-анализ / Л. Аверьянов, 2009. – Режим доступу: [http://www.sbiblio.com/biblio/archive/averjanov\\_kontent/](http://www.sbiblio.com/biblio/archive/averjanov_kontent/). – Назва з тит. екрану. 2. Большакова Е. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учеб.пособие / Е. Большакова, Э. Клышинский, Д. Ландэ, А. Носков, О. Пескова, Е. Ягунова. – М.: МИЭМ, 2011. – 272 с. 3. Валгина Н. Теория текста: учебное пособие / Н. Валгина. – М: Логос, 2003. – 280 с. – Режим доступу: <http://evartist.narod.ru/text14/01.htm>. – Назва з титул. екрану. 4. Григорьев С. Проведение контент-анализа / С. Григорьев. – Режим доступу: <http://www.psyfactor.org/lib/k-a2.htm>. – Назва з тит. екрану. 5. Дмитриев И. Контент-анализ: сущность, задачи, процедуры / И. Дмитриев. – Режим доступу: <http://www.psyfactor.org/lib/k-a.htm>. – Назва з тит. екрану. 6. Клифтон Б. Google Analytics: анализ посещаемости веб-сайтов / Б. Клифтон. – М: ООО “И.Д. Вильямс”, 2009. – 400 с. 7. Леліков Г. Моніторинг діяльності органів виконавчої влади із застосуванням комп'ютерної системи контент-аналізу електронних ЗМІ / Г. Леліков, В. Сороко, О. Григор'єв, Д. Ланде // Вісник державної Служби України, N 2'2002. – Режим доступу: <http://www.infostream.ua/infostream/publ/guds/index.shtml>. – Назва з тит. екрану. 8. Ландэ Д. Интернетика. Навигация в сложных сетях: модели и алгоритмы / Д. Ландэ, А. Снарский, И. Безсуднов // Москва-2009. – Режим доступу: [http://webground.su/services.php?param=book&part=internetica\\_content.htm](http://webground.su/services.php?param=book&part=internetica_content.htm). – Назва з тит. екрана. 9. Ландэ Д. Основы моделирования и оценки электронных информационных потоков: монографія / Д. Ландэ, В. Фурашев, С. Брайчевский, О. Григорьев. – К.: ТОВ “Інжиніринг”, 2006. – 348 с. 10. Ландэ Д. Феномены современных информа-ционных потоков / Д. Ландэ, А. Литвин // Журнал “Сети и бизнес” N 1, 2001, ИЦ “ЭЛВИСТИ”. – Режим доступу: <http://www.infostream.ua/publ/content/>. – Назва з тит. екрана. 11. Манаев О. Контент-анализ – описание метода. Контент-анализ как метод исследования / О. Манаев. – Режим доступу: <http://www.psyfactor.org/lib/kontent.htm>. – Назва з тит. екрана. 12. Писаревська Т. Інформаційні системи і технології в управлінні трудовими ресурсами / Т. Писаревська // 2-ге вид., перероб. і доп. – К.: КНЕУ, 2000. – 279 с. 13. Попов М. Визуализация семантической структуры и реферирование текстов на естественном языке / М. Попов, А. Заболева-Зотова, С. Фоменков // Волгоградский Государст-венный Университет. – Режим доступу: <http://www.dialog-21.ru/Archive/2003/Popov.htm>. – Назва з тит. екрана. 14. Ситник В. Основы інформаційних систем / В. Ситник, Т. Писаревська, Н. Єрьоміна, О. Краєва // За ред. В. Ситника. – 2-е вид., перероб. і доп. – К.: КНЕУ, 2001. – 420 с. 15. Федорчук А. Контент-мониторинг информационных потоков / А. Федорчук // Б-ки нац. акад. наук: пробл. функционирования, тенденции развития. – К., 2005. – Вып. 3. – Режим доступу: <http://www.nbuv.gov.ua/articles/2005/05fagmip.html>. – Назва з тит. екрану. 16. Харченко К. Компьютерный контент-анализ текстовой информации / К. Харченко. – Режим доступу: <http://leveltar.narod.ru/cta/teor.html>. – Назва з титул. екрана.